

Article

Gráfico de control T^2 Hotelling para variables cualitativas

Wilson Rojas-Preciado^{1,2,*} , Mauricio J. Rojas-Campuzano³ , Purificación Galindo-Villardón² , Omar Ruiz-Barzola³ 

¹ Universidad Técnica de Machala - Machala, Ecuador; wrojas@utmachala.edu.ec

² Universidad de Salamanca Salamanca, España; wrojas@usal.es; pgalindo@usal.es

³ Escuela Superior Politécnica del Litoral Gayaquil, Ecuador; maujroja@espol.edu.ec; oruiz@espol.edu.ec

* Correspondence: wrojas@utmachala.edu.ec; Tel.: +593-992-83-3719

† Current address: Updated affiliation

Version March 21, 2023 submitted to Mathematics



Simple Summary: A Simple summary goes here.

Abstract: La literatura científica es abundante en lo referente a gráficos de control en entornos multivariantes para datos numéricos y mixtos, sin embargo, para datos cualitativos hay pocas publicaciones. Las variables cualitativas aportan valiosa información de procesos en diversos contextos industriales, productivos, sociales. Los procesos educativos no son una excepción, tienen múltiples variables asociadas a estudiantes, profesores e instituciones. Cuando hay muchas variables se corre el riesgo de tomar información redundante o excesiva, luego, es viable la aplicación de métodos multivariantes de reducción de dimensiones para quedarse con pocas variables ficticias, combinación de las reales, que sintetizan la mayor parte de la información. En este contexto se presenta el gráfico de control T2Qv, una técnica de control estadístico de procesos multivariantes que realiza un análisis de datos cualitativos mediante Análisis de correspondencias múltiples (MCA), Análisis Factorial Múltiple y el gráfico T^2 de Hotelling. La interpretación de los puntos fuera de control se realiza comparando los gráficos MCA y analizando la distancia X^2 entre las categorías de la tabla concatenada y las que representan puntos fuera de control. El análisis de sensibilidad determinó que el gráfico de control T2Qv tiene un buen rendimiento cuando trabaja con altas dimensiones. Para probar la metodología se hizo un análisis con datos simulados y otro con datos reales relacionados con la educación superior. Para facilitar la difusión y aplicación de la propuesta, se desarrolló un paquete computacional reproducible en R, denominado T2Qv y disponible en The Comprehensive R Archive Network (CRAN).

Keywords: Multivariate; Statistical Process Control; Qualitative; Control Charts; R; T2 Hotelling; Superior Education.

1. Introduction

El control estadístico juega un rol muy importante en la mejora continua de los procesos y dentro de éste, los gráficos de control, los cuales ayudan a monitorizar los procesos, han sido extensamente utilizados desde su creación por Walter Shewhart [1] hasta nuestros días. A partir de los gráficos univariantes se han desarrollado un sinnúmero de propuestas, las cuales incorporaron la opción de monitorizar varias variables a la vez [2,3], abriendo con ello el control estadístico de procesos multivariantes (CEPM).

Las opciones más conocidas en el CEPm son: El gráfico de control T2 de Hotelling [4], el cual se podría considerar la versión multivariante del gráfico de medias de Shewhart; el MEWMA [5], el cual

es la versión multivariante del gráfico de medias ponderadas EWMA [6]; o el MCUSUM [7], el cual es la versión multivariante del gráfico de control de sumas acumuladas CUSUM [8].

A estos gráficos de control multivariante se le han realizado varias mejoras tales como: su optimización, determinando de forma analítica [9–11] o heurística los valores óptimos de sus parámetros [12]; otra propuesta es la de trabajar sin distribuciones probabilísticas o versiones no paramétricas [13–15], para procesos continuos o por lotes [2].

Todos estos gráficos de control multivariantes tienen un enfoque cuantitativo; es decir, las variables monitorizadas son esencialmente cuantitativas, ya sean discretas o continuas. Para ello, los diferentes autores, inicialmente se valieron de la distancia de Mahalanobis [16]. Posteriormente, para el análisis de una combinación de variables continuas y categóricas se desarrolló el gráfico basado en la distancia de Gower [17]. Sin embargo, el abordaje de problemas como la alta correlación entre características y en presencia de datos mixtos, requirió la incorporación de técnicas estadísticas multivariantes clásicas, como Análisis de Componentes Principales [18], Métodos Biplot [19,20], Análisis de correspondencias [21], STATIS [22–24], Coordenadas paralelas [25], Análisis de conglomerados [26].

Dentro de las aportaciones referentes a los gráficos de control que incorporan técnicas multivariantes destacan el gráfico basado en STATIS para el monitoreo de procesos por lotes en entornos no paramétricos [27]; los diagramas de bolsa robustos que utilizan STATIS Dual y Coordenadas paralelas [28]; el Gráfico de control multivariante PCA para datos mixtos, que aplica una combinación de Análisis de Componentes Principales y Análisis de Correspondencias Múltiples [29]; el gráfico de control de peso de novedad sensible a la densidad (DNW) utiliza el algoritmo k -vecino más cercano (kNN) [14]; el gráfico basado en Kernel PCA Mix [30,31]; el Gráfico T^2 basado en la combinación de PCA para datos continuos y cualitativos con detección de datos atípicos [32]; los gráficos de control basados en PCA para entornos no paramétricos [33,34].

No obstante, las aportaciones al desarrollo de gráficos de control multivariante para variables cualitativas no han sido numerosas. En este campo las propuestas se han desarrollado alrededor del análisis de variables que siguen una distribución Poisson y el análisis de variables multinomiales. La primera propuesta fue la de Holgate [35], quien presentó un trabajo sobre la distribución Poisson bivalente para variables correlacionadas. Este modelo fue tomado como insumo en las investigaciones de autores como Chiu and Kuo [36], Lee and Costa [37], Laungrungrong *et al.* [38], Epprecht *et al.* [39].

Otra propuesta destacada es la de Lu [40], quien desarrolló un gráfico de control tipo Shewhart para procesos multivariantes con variables cualitativas, cuando la característica de calidad asume valores binarios, al cual denominaron gráfico np multivariante (MNP). Ya en el contexto multinomial, Ranjan-Mukhopadhyay [41] planteó un gráfico de control multivariante utilizando el estadístico D^2 de Mahalanobis para atributos que siguen una distribución multinomial. Además, de una propuesta para procesos multinomiales bajo el enfoque difuso [42]; Taleb [43] introdujo gráficos de control para el monitoreo de procesos multivariados con datos lingüísticos multidimensionales, basados en dos procedimientos: la teoría de la probabilidad y la teoría difusa; Pastuizaca-Fernández *et al.* [44] presentaron un gráfico de control multivariante multinomial T2 con un enfoque difuso.

Salto Segura *et al.* [45] aseguran que las herramientas de control de la calidad se pueden considerar no solo para monitorizar procesos industriales sino también procesos relacionados con la educación, por ejemplo, la evaluación del desempeño estudiantil. Estos autores aplicaron el concepto de profundidad, que transforma una observación multivariante a un índice univariante, el cual es susceptible de monitorizar en una carta de control y para esto utilizaron la carta r , además utilizaron clúster medio para establecer umbrales que faciliten la conformación de grupos y establecer perfiles de estudiantes mediante medidas descriptivas.

En el estudio de los procesos que se desarrollan en el entorno social se maneja con mucha frecuencia variables cualitativas. No es que estén ausentes los datos cuantitativos, sino que, en las bases de datos que se utilizan para estos análisis, abundan las variables cualitativas nominales y ordinales, a veces, sobre las de tipo numérico.

López [46] señala que al observar muchas variables sobre una muestra es presumible que una parte de la información recogida pueda ser redundante o que sea excesiva, en cuyo caso los métodos multivariantes de reducción de la dimensión tratan de eliminarla combinando muchas variables observadas para quedarse con pocas variables ficticias que, aunque no observadas, sean combinación de las reales y sintetizan la mayor parte de la información contenida en sus datos. En este caso se deberá tener en cuenta el tipo de variables que maneja. Si son variables cuantitativas las técnicas que le permiten este tratamiento pueden ser el Análisis de componentes principales [47,48], el Análisis factorial [49–51], mientras que, si se trata de variables cualitativas, es recomendable la aplicación de un Análisis de correspondencias múltiple, Análisis de homogeneidad o un Análisis de Escalamiento multidimensional.

En el control estadístico de procesos, los aportes al desarrollo de gráficos de control para variables cualitativas todavía son incipientes, las pocas publicaciones se orientan al análisis de características de la calidad en procesos industriales, pero no a procesos sociales. Al analizar los procedimientos publicados por los autores citados en este estudio, se detectan limitaciones que podrían restringir su aplicación, por ejemplo, el análisis de pocas características de la calidad, el uso de muestras constituidas por elementos individuales en vez de grupos, la dificultad de trabajar con muchas categorías de forma simultánea. Surge, entonces, la necesidad de un gráfico de control para la representación de p variables cualitativas, que pueda trabajar con múltiples categorías nominales y ordinales y que facilite la identificación de las causas que pueden llevar al proceso a un estado fuera de control y que pueda ser aplicado en procesos sociales.

El presente artículo atiende las limitaciones antes mencionadas en cuanto a gráficos de control para variables cualitativas y su aplicación en entornos sociales. Por tal motivo, su objetivo es desarrollar un gráfico de control para variables cualitativas mediante el uso de metodologías estadísticas multivariantes, para que se contribuya a la diversificación de técnicas en la fase I del control estadístico de procesos.

Este artículo está organizado de la siguiente manera: la Introducción, que establece los antecedentes conceptuales y referenciales de los gráficos de control multivariantes aplicados a variables cualitativas; la sección 2, materiales y métodos, que detalla el procedimiento que se siguió en el desarrollo del gráfico de control propuesto; la sección 3 describe al complemento computacional que facilita la aplicación de esta metodología; la sección 4 muestra los resultados mediante el análisis de datos simulados; la sección 5 corresponde al análisis de sensibilidad que relaciona el número de dimensiones analizadas versus la confiabilidad de los resultados. La sección 6 presenta la discusión mediante un análisis comparativo entre el gráfico de control T2Qv y las propuestas de otros autores. Finalmente, la sección 7 establece las conclusiones.

2. Metodología

2.1. Notación

La tabla 1 contiene elementos, representación y ejemplos de la manera como se presentan los elementos algebraicos abordados en la metodología.

Elementos	Representación	Ejemplo
Escalares	Letras en minúscula.	v, λ
Vectores	Letras en minúscula y en negrita.	\mathbf{v}, \mathbf{u}
Matrices	Letras en mayúscula y en negrita.	\mathbf{V}, \mathbf{X}
Matrices de tres vías (Cubos de datos)	Letras con doble trazo en mayúscula.	\mathbb{C}, \mathbb{X}

Table 1. Elementos algebraicos

A lo largo del artículo se utilizarán letras para hacer referencia a parámetros necesarios, se los enuncia a continuación en la tabla 2:

Letra	Significado	Especificación
p	Número de dimensiones	
K	Número total de tablas (Especifica la profundidad del cubo de datos)	
k	Índice de tabla	$k=1,2,\dots,K$
T	Índice de matriz transpuesta	X^T
n	Tamaño muestral de las k tablas	

Table 2. Notación

2.2. Análisis de Correspondencias Múltiples (MCA)

Dado que se trabaja con variables cualitativas se aplica Análisis de Correspondencias [21] analizando la similaridad entre categorías [46] basada en la distancia χ^2 , siendo este un análisis similar al de componentes principales.

En este caso no se aplica el enfoque francés [52], si no el anglosajón, donde el MCA se denomina Análisis de Homogeneidad o Escalamiento Dual, haciendo uso de la tabla de Burt [21] y partiendo de una matriz de datos con p variables cualitativas, cada una con h categorías ($h > 1$). La matriz compuesta por las n filas u observaciones y las p columnas o variables, donde cada celda contenga una de las categorías antes mencionadas, es equivalente a la matriz disyuntiva Z , que desglosa las variables en cada una de sus modalidades y registra la ocurrencia de eventos de forma binaria [21].

La tabla de Burt viene dada por:

$$B = Z'Z \quad (1)$$

La matriz B en 1 está formada por las frecuencias absolutas, éstas se transforman en frecuencias relativas, dividiendo los valores de la matriz por la frecuencia total, dando lugar a la matriz P .

Se obtienen los vectores de *Masas de fila* (mf) y *columna* (mc), a través de las marginales de las filas y de las columnas de la matriz P , respectivamente.

Se obtiene la matriz de residuos estandarizados S .

$$S = D_{\text{fila}}^{-\frac{1}{2}}(P - mf \ mc')D_{\text{columna}}^{-\frac{1}{2}} \quad (2)$$

donde D_{fila} es una matriz diagonal que contiene las masas de las filas y D_{columna} es una matriz diagonal que contiene las masas de las columnas.

Se aplica descomposición singular (SVD) a la matriz S (Ecuación 2):

$$S = UDV' \quad (3)$$

donde U y V son matrices ortogonales y D es una matriz diagonal que contiene los valores singulares.

Luego se obtienen las coordenadas estandarizadas, para lo cual se aplican las ecuaciones 4 y 5.

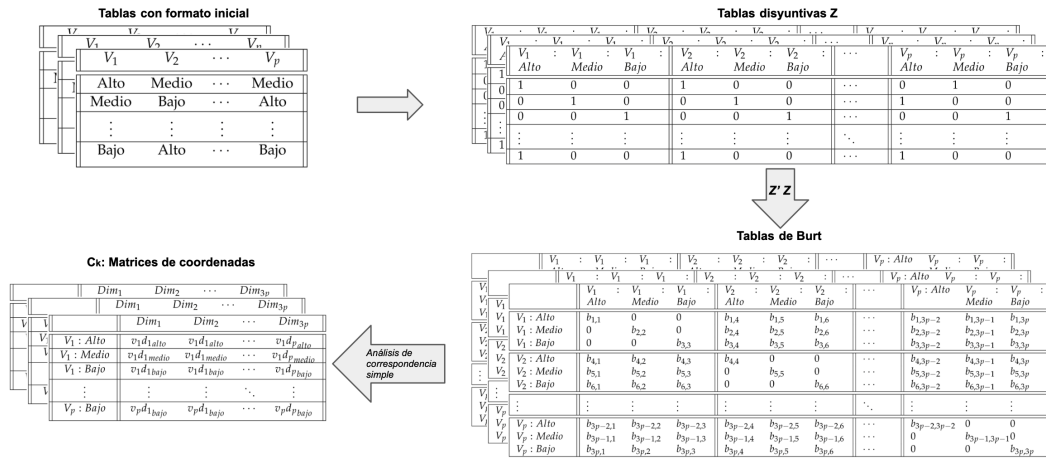
$$X = D_{\text{fila}}^{-\frac{1}{2}}U \quad (4)$$

$$Y = D_{\text{columna}}^{-\frac{1}{2}}V \quad (5)$$

2.2.1. Generalización a k tablas

Si se tienen K tablas, de igual estructura y compuestas por variables cualitativas, a cada una de las K tablas se aplica lo descrito en 2.2, obteniéndose el conjunto de K tablas con formato inicial ??.

Se denomina C a cada una de las K tablas de coordenadas obtenidas en el paso anterior. Con la finalidad de detectar la magnitud de las variables latentes, se aplica valor absoluto a los elementos de

Figure 1. Procedimiento del MCA para k tablas

la matriz C_k ($k = 1, \dots, K$). De esta forma se obtiene un conjunto de k tablas de coordenadas (cargas), cuyas filas corresponden a las variables observadas y las columnas a las variables latentes.

2.2.2. Normalización de tablas

A las K tablas C se le aplica la normalización [53] del Análisis factorial múltiple (MFA).

Sea λ_1^k el primer valor propio obtenido de la descomposición singular de la k -ésima tabla C . Se normaliza la tabla multiplicándola por $1/\lambda_1^k$. Con esto se obtiene la tabla C' , que corresponde a la tabla de coordenadas normalizadas.

Individualmente, para el caso de la matriz k , se tendría la siguiente expresión.

$$C'_k = \frac{1}{\lambda_1^k} C_k \quad (6)$$

Hasta este punto se tiene un conjunto de matrices de coordenadas normalizadas, cuyas filas contienen las variables observadas y las columnas, las variables latentes.

La expresión de la ecuación 6 aplicada a k tablas se representa en la figura ??, que muestra el esquema de preparación de las tablas, previo a la obtención de vectores de centralidad usados por el gráfico de control multivariante.

Unificando las K tablas normalizadas C' en una sola, se tiene la matriz C' , denominada Matriz Concatenada. Esta contiene todos los elementos de las K tablas normalizadas.

$$C' = [C'_1 | C'_2 | \dots | C'_K]^T \quad (7)$$

La normalización que realiza el MFA se encarga de ponderar las K tablas, con el objetivo de evitar alguna descompensación al momento de realizar el análisis conjunto de las tablas.

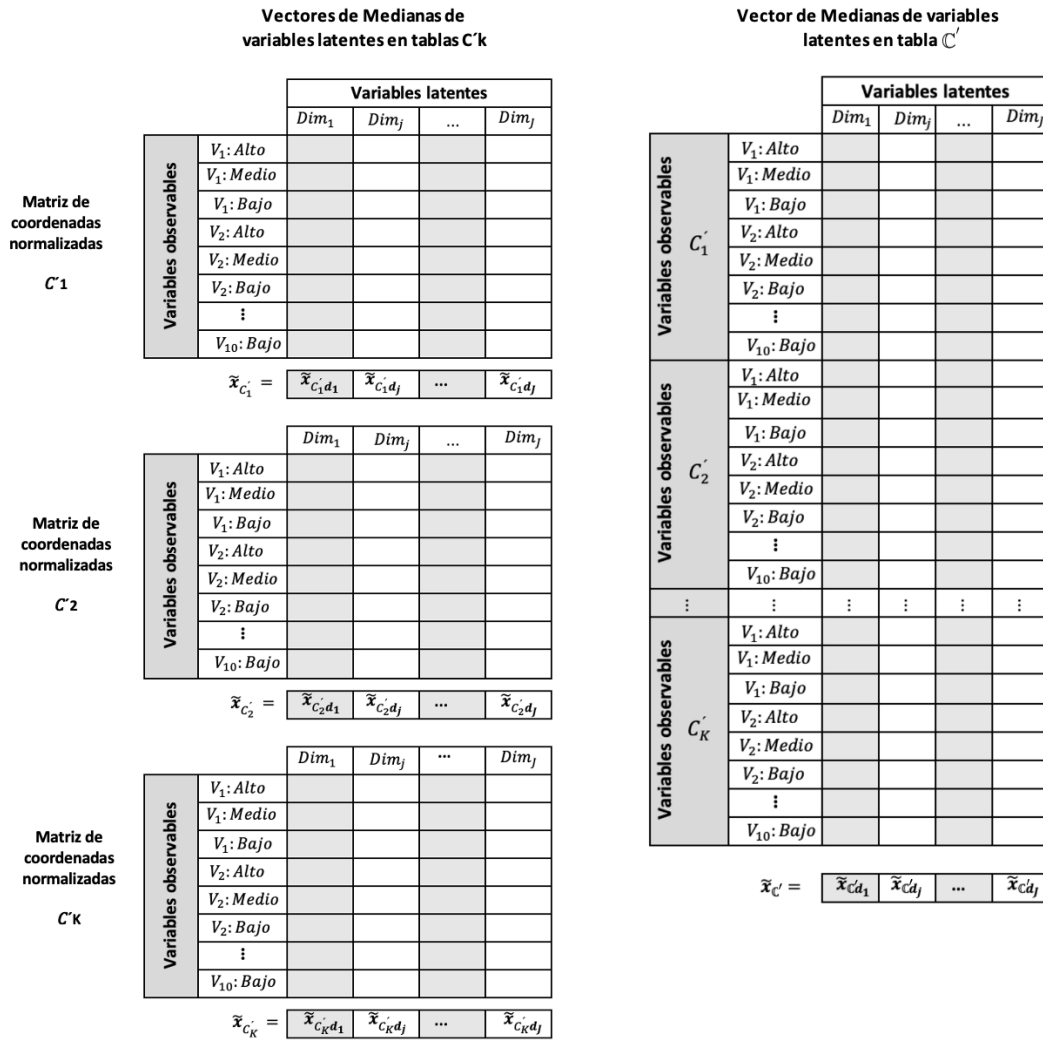
A partir de las matrices C' y C'_k , se obtiene los vectores de mediana, tal como se muestra en la figura 2. El vector $\tilde{x}_{C'_k}$ explicará el comportamiento central de la tabla k y el vector $\tilde{x}_{C'}$ explicará el comportamiento de la matriz concatenada.

2.3. Gráfico de control $T2Q\bar{v}$

2.3.1. Obtención del gráfico de control

Para definir el gráfico de control T^2 Hotelling se deben tomar las siguientes consideraciones:

- La tabla C' (Ecuación 7) se denomina Concatenada, sirve como referente para el escenario *bajo control* en la fase I del control del proceso.

**Figure 2.** Esquema de obtención de vectores de medianas

- El estadístico T^2 Hotelling normalmente se calcula con los vectores de media y la matriz de covarianzas del proceso bajo control; la propuesta de esta investigación es adoptar conceptos de robustez, utilizando el vector de medianas en vez de el vector de medias, en virtud de que a las medianas no les afectan los valores atípicos.
- De la matriz concatenada C' se obtiene \tilde{x}_0 (Vector de medianas de la matriz concatenada) y S_0 (Matriz de covarianzas de la matriz concatenada).
- Cada matriz C'_k tiene el mismo número de columnas.
- El vector de medias \tilde{x}_k está atado a la tabla C'_k , es decir, el gráfico de control estará en función de las diferencias entre las matrices C'_k y la matriz concatenada C' .
- Las matrices C'_k siguen una distribución normal multivariante con vector de mediana \tilde{x}_k y matriz de covarianzas S_k .

El estadístico T^2 viene dado por:

$$T^2 = n(\mu_k - \mu_0)' \Sigma_0^{-1} (\mu_k - \mu_0) \quad (8)$$

Tomando en cuenta las consideraciones previas, se obtiene el estadístico T_{med}^2

$$T_{med}^2 = n(\tilde{x}_k - \tilde{x}_0)' \Sigma_0^{-1} (\tilde{x}_k - \tilde{x}_0) \quad (9)$$

Se sabe que, bajo control, el T^2 se distribuye como una Chi-cuadrado con p grados de libertad χ_p^2 . En este caso se puede aplicar este principio, ya que se utiliza la matriz concatenada (\mathbb{C}'), que representa al escenario bajo control.

Dado que este gráfico de control está basado en distancias de Mahalanobis ponderadas, sólo tiene límite de control superior. Este viene dado por la ecuación 10

$$UCL = \chi_{\alpha, p}^2 \quad (10)$$

donde p es el número de dimensiones y α es la significancia predeterminada, se considera $\alpha = 0.0027$.

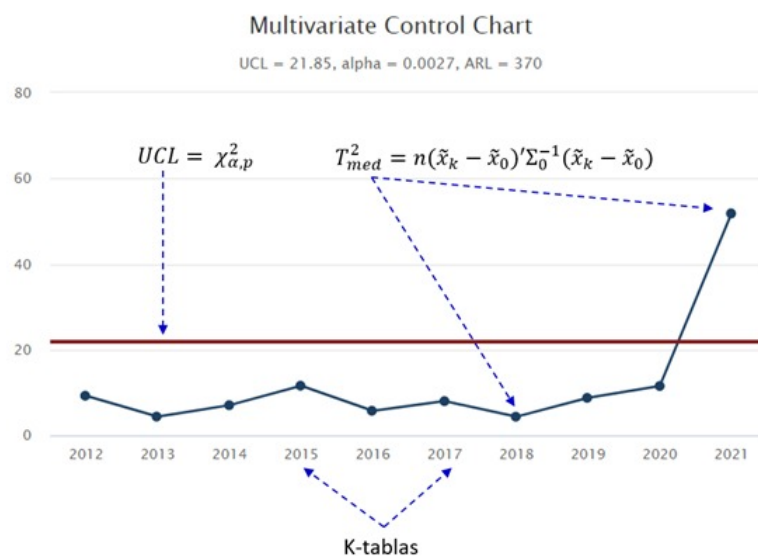


Figure 3. Gráfico T2Qv

2.3.2. Interpretación de puntos fuera de control

El gráfico multivariante para variables cualitativas, T2Qv, es capaz de señalar que el proceso salió de control, pero no permite reconocer las causas por las que ocurrió esto. Cada punto representado en el gráfico representa a una tabla (muestra), constituida por un grupo de individuos (observaciones) y p variables que pueden tener muchas categorías, algunas de éstas pueden mostrar un comportamiento anómalo. Por consiguiente, es necesario analizar con detenimiento que está pasando con los datos de las tablas reportadas para identificar la(s) variable(s) que generaron que el proceso se salga de control.

Este análisis se realiza comparando la ubicación de los puntos que representan las categorías de las variables en el MCA de la tabla concatenada y la ubicación de los puntos en los gráficos MCA de cada tabla reportada como fuera de control. Las categorías que están incidiendo en el estado fuera de control son aquellas que muestran diferencias notorias en su ubicación al comparar ambas tablas. Para cuantificar la magnitud de dichas diferencias o del comportamiento anómalo de estas categorías se calcula las distancias Chi-cuadrado entre las masas de las columnas de la tabla reportada como fuera de control y las columnas de la tabla concatenada, tomada esta última como referente. Mientras mayor es el valor del estadístico, mayor es su incidencia en el desplazamiento de la centralidad del proceso que, finalmente, pueden llevarlo a un estado fuera de control.

3. Complemento computacional

Para facilitar la difusión y aplicación del método propuesto, se ha desarrollado un paquete reproducible en R. El paquete **T2Qv** [54] realiza el análisis de control de K tablas por medio de gráficos

de control multivariantes para variables cualitativas, utilizando los fundamentos teóricos del análisis de correspondencias múltiples y el análisis factorial múltiple, así como la idea conceptual del STATIS.

Los gráficos se pueden mostrar de forma plana o interactiva, de la misma manera todas las salidas se pueden mostrar en un panel interactivo de Shiny y sus resultados gráficos y numéricos pueden ser exportados.

3.1. Descripción del paquete T2Qv

El paquete estadístico T2Qv realiza Análisis de Correspondencias Múltiples a las tablas originales (\mathbf{T}_k), generando matrices de variables latentes (\mathbf{C}_k) cuyas coordenadas se someten a un proceso de normalización, multiplicándolas por $1/\lambda_1^k$. Las matrices de coordenadas normalizadas (\mathbf{T}'_k) se ordenan una debajo de otra, para conformar una tabla concatenada (\mathbf{C}'), de la que se extrae su vector de medianas $\mathbf{x}_{C'}$ así como los vectores de medianas de cada matriz $\mathbf{x}_{C'_k}$ que la conforman. Con estos vectores se obtienen los estadísticos $T_{med}^2 = n(\tilde{x}_k - \tilde{x}_0)' \Sigma_0^{-1} (\tilde{x}_k - \tilde{x}_0)$ para cada una de las tablas analizadas, los que se representan como puntos en el gráfico de control T2Qv. Los puntos que en el gráfico se ubiquen fuera del límite ($UCL = \chi_{\alpha,p}^2$) son reportados fuera de control.

El paquete estadístico T2Qv permite la interpretación del comportamiento anómalo de los puntos fuera de control a través de la comparación de los gráficos MCA de una tabla \mathbf{TC}_k , que resulta de concatenar las matrices iniciales, y cada tabla inicial \mathbf{T}'_k . El paquete permite la selección de las \mathbf{T}'_k tablas, de manera que el investigador pueda enfocar su análisis en las que se identifiquen como fuera de control.

Además, el paquete T2Qv genera un gráfico interactivo de barras que representa las distancias χ^2 entre las masas de columnas de las variables de la tabla \mathbf{TC}_k y la tabla \mathbf{T}'_k . Las barras que denotan mayor altura identifican a las variables que están provocando, con mayor fuerza, la salida de control de la k -ésima tabla. Este gráfico interactivo incluye, mediante un gráfico circular anidado, una representación de la distribución de las categorías de la variable observada, correspondiente a la k -ésima tabla, así como un gráfico circular de la distribución de las categorías de la tabla concatenada (\mathbf{TC}_k), lo que facilita la identificación de los cambios en la distribución de las categorías.

De esta manera el paquete T2Qv consolida la metodología propuesta en esta investigación y permite la explicación de cuándo y por qué el proceso salió de control.

Las funciones que incluye el paquete y su descripción se enuncian en la tabla 3.

Función	Descripción
T2 qualitative	Multivariate control chart T2 Hotelling applicable for qualitative variables.
MCAconcatenated	Multiple correspondence analysis applied to a concatenated table.
MCApoint	Multiple correspondence analysis applied to a specific table.
ChiSq variable	Contains Chi square distance between the column masses of the table specified in PointTable and the concatenated table. It allows to identify which mode is responsible for the anomaly in the table in which it is located.
Full Panel	A shiny panel complete with the multivariate control chart for qualitative variables, the two MCA charts and the modality distance table. Within the dashboard, arguments such as type I error and dimensionality can be modified.

Table 3. Funciones del paquete T2Qv

3.2. Disponibilidad

El paquete está disponible en el repositorio oficial de R, The Comprehensive R Archive Network (CRAN), la descarga se la puede realizar de la siguiente forma:

```
install.packages("T2Qv")
```


4. Resultados

Con la intención de probar la metodología propuesta en el gráfico de control T^2 de Hotelling para variables cualitativas, se hizo un análisis con datos simulados y otro con datos reales aplicados al contexto de la educación superior. Los resultados se obtienen de la aplicación del paquete T2Qv.

4.1. Resultados con datos simulados

4.1.1. Generación de datos simulados

Para este estudio se generó una base de datos simulados, a la que se denominó *Datak10Contaminated*. Consta de 10 tablas, cada una de ellas está constituida por 100 filas (observaciones) y 11 columnas, de las cuales, las 10 primeras corresponden a las variables analizadas ($V1, V2, \dots, V10$), mismas que contienen 3 categorías (Alto, Medio y Bajo), mientras que, la columna 11, denominada *GroupLetter*, contiene el factor de clasificación de los grupos. Para su identificación, las tablas han sido denominadas con las letras del alfabeto, desde la *a* hasta la *j*. La tabla *j* tiene una distribución distinta de la que tienen las otras nueve. Las 9 primeras tablas tienen sus 10 variables con la siguiente distribución:

$$u \sim U[0,1]$$

$$t_{1,\dots,9} = \begin{cases} \text{Bajo} & \text{si} & u \leq 1/3 \\ \text{Medio} & \text{si} & 1/3 < u < 2/3 \\ \text{Alto} & \text{si} & u \geq 2/3 \end{cases}$$

La tabla *j* o tabla 10, en todas sus 10 variables, sigue la distribución presentada a continuación:

$$u \sim U[0,1]$$

$$t_j = \begin{cases} \text{Bajo} & \text{si} & u \leq 1/5 \\ \text{Medio} & \text{si} & 1/5 < u < 2/6 \\ \text{Alto} & \text{si} & u \geq 2/6 \end{cases}$$

La base de datos se presenta en el formato establecido en la tabla 4.

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	GroupLetter
Low	Medium	Medium	High	High	High	Low	Medium	Medium	Medium	a
Low	Low	High	Low	Medium	High	High	High	Low	High	a
High	Medium	High	Low	High	Medium	Medium	High	Medium	Low	a
Medium	Medium	Low	High	Low	Medium	High	Low	Low	High	a
Low	Low	Low	High	Low	High	High	High	Medium	Medium	a
High	High	Medium	Low	High	Low	Medium	Medium	High	Low	a
High	High	Low	Low	Low	Medium	High	Medium	Medium	High	a
Medium	Medium	High	Medium	Medium	High	Medium	High	High	High	a
Low	Low	Low	Medium	High	Medium	Low	Medium	Low	Low	a
Medium	Medium	Medium	High	Low	Medium	High	Low	High	Medium	a

Table 4. Sección de la base de datos *Datak10Contaminated*.

Para verificar la diferencia entre las distribuciones de la tabla 10 y las demás, se calculó el promedio de las frecuencias relativas en las tres categorías, desde la tabla *a* hasta la *i*, para las 10 variables (Anexo 1), luego se calculó el promedio de las frecuencias relativas medias de las 10 variables, el resultado

permite comparar la distribución de las categorías de la tabla *Datak10Contaminated* con la distribución teórica uniforme, como se observa en la tabla 5.

Categorías	Teórica uniforme	Promedio Tablas a, b, \dots, i	Promedio Tabla j
High	0.333	0.340	0.724
Medium	0.333	0.336	0.092
Low	0.333	0.324	0.184

Table 5. Comparación de la distribución de las categorías de la tabla *Datak10Contaminated* con la distribución teórica uniforme.

Se aplicaron las respectivas pruebas Chi cuadrado de bondad de ajuste para confirmar la distribución de los datos generados, así como la comparación de la tabla j con las demás tablas, confirmándose las diferencias significativas entre las distribuciones (p -valor < 0.05), como se observa en el anexo 2.

4.1.2. Aplicación del paquete T2Qv con datos simulados

El primer resultado es el gráfico del Análisis de Correspondencias Múltiples (MCA) aplicado a la tabla concatenada (Figura 4). Esta tabla es considerada el referente visual para el escenario bajo control para el análisis posterior de las tablas que sean reportadas como puntos fuera de control en el gráfico T2Qv.

El MCA reporta una inercia total del 63.35%, la dimensión 1 representa al 53.64% de la información, mientras que la dimensión 2, al 9.71%. Los puntos del gráfico representan a las observaciones de cada una de las 10 variables en sus tres niveles: *High*, *Medium* y *Low*. En el gráfico del MCA, las observaciones que se ubiquen en el centro del gráfico representan a las categorías que se presentan con mayor frecuencia, mientras que las más alejadas del centro son las que pocas veces aparecen, los casos raros. En este sentido, en la tabla concatenada no hay observaciones ubicadas en el centro del gráfico, sino que están repartidas en grupos, rodeando el centro, lo que se explica por la distribución uniforme de las categorías de las variables en la mayoría de las tablas, ninguna prevalece.

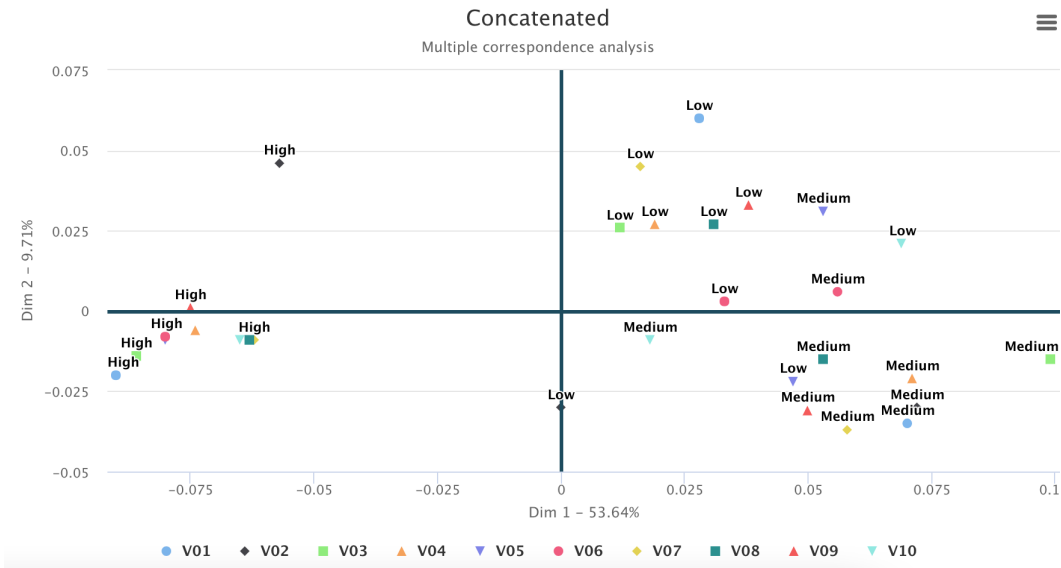


Figure 4. Análisis de correspondencias múltiples aplicado a la tabla concatenada.

Otro resultado es el MCA aplicado a una tabla específica. En este punto, uno de los argumentos que se debe tener en cuenta es la selección de la tabla con la que se realizará el análisis.

Al comparar los gráficos, se puede observar que la tabla del punto b (5) muestra diferencias con la tabla del estado bajo control (4); sin embargo, las diferencias no llegan a ser significativas, por lo que no genera una señal de fuera de control (6, punto b).

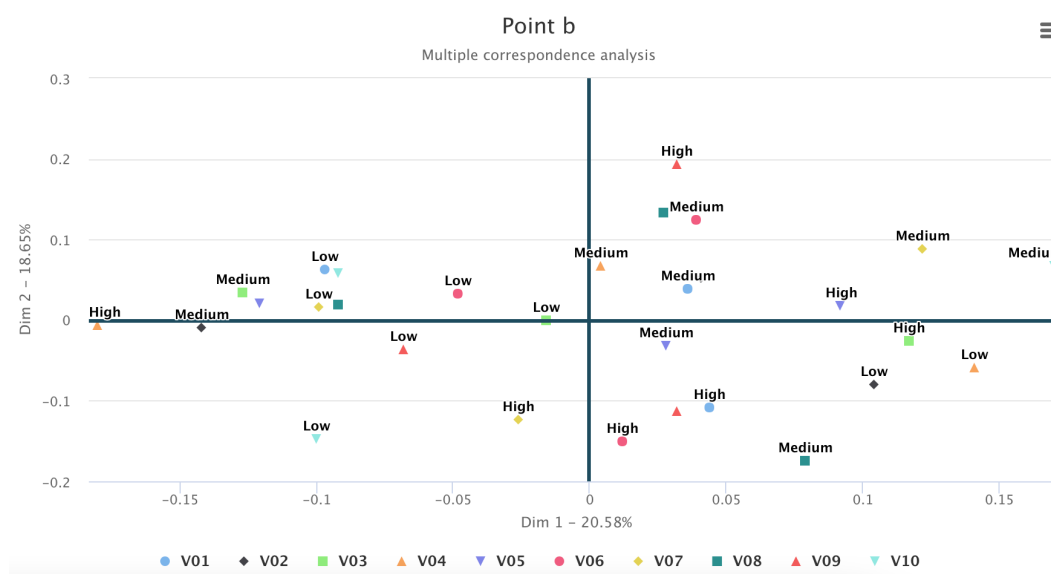


Figure 5. Análisis de correspondencias múltiples aplicado a la tabla b.

La figura 5 representa el gráfico del MCA de la tabla b, correspondiente a un momento específico del proceso monitorizado. Este gráfico, en sus dos dimensiones, representa al 39.23% de la información. Es notorio que las observaciones en sus niveles *alto*, *medio* y *bajo* están distribuidas de forma aleatoria en todos los cuadrantes del gráfico, no se puede precisar un patrón específico de agrupación.

Lo mismo se puede decir de los puntos representados en cualquiera de las otras tablas porque comparten la misma distribución, exceptuando la tabla j, que fue diseñada con una distribución diferente. No obstante, el uso del MCA de las figuras 5 y 4 todavía no permite detectar si el proceso está o no en control. La identificación de puntos fuera de control se puede realizar mediante la representación gráfica del estadístico T_{med}^2 , como se observa en la figura 6.

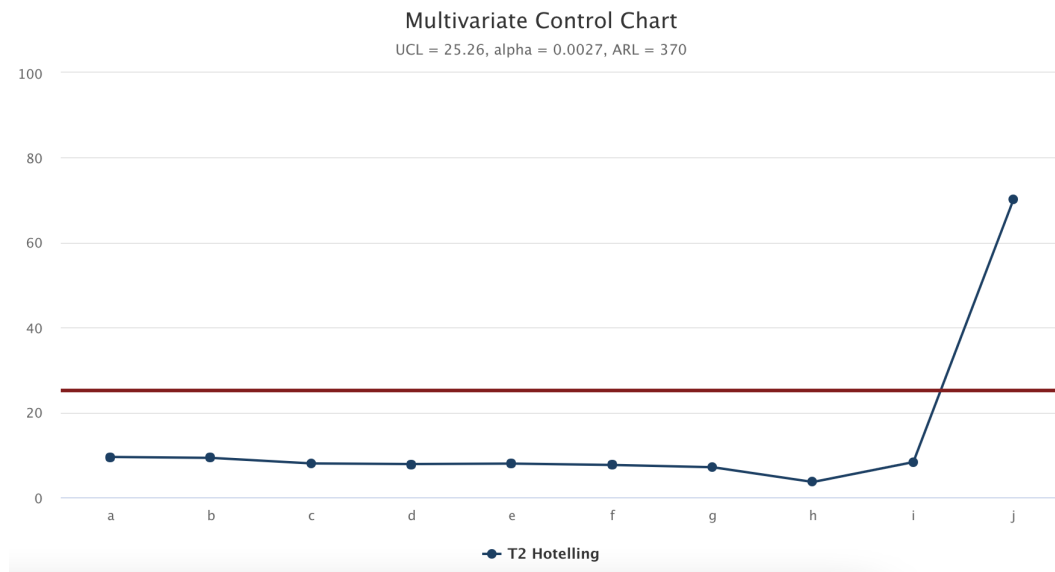


Figure 6. Gráfico de control multivariante T2 Hotelling aplicable a variables cualitativas, *Datak10Contaminated*.

La figura 6 presenta al gráfico de control T2Qv, basado en el estadístico T2 de Hotelling ajustado (T_{med}^2), aplicado a la detección de anomalías en cualquiera de las K tablas analizadas. Cada una de las tablas está representada por los puntos en el gráfico. Se observa una línea horizontal que representa al límite de control superior (UCL). El límite de control inferior (LCL) es igual a cero.

Se observa que el punto que representa a la tabla j se ubica por encima del límite de control superior, lo que quiere decir que se lo ha identificado como un valor fuera de control. Por consiguiente, es necesario analizar con detenimiento qué está pasando con los datos de la tabla reportada, comparándolos con los de la tabla concatenada, a fin de identificar las causas de la variación y tomar las acciones pertinentes. Para hacer un análisis del punto fuera de control se realiza un gráfico del MCA de la tabla j y se lo compara con el gráfico similar de la tabla concatenada, como se presenta en la figura 7.

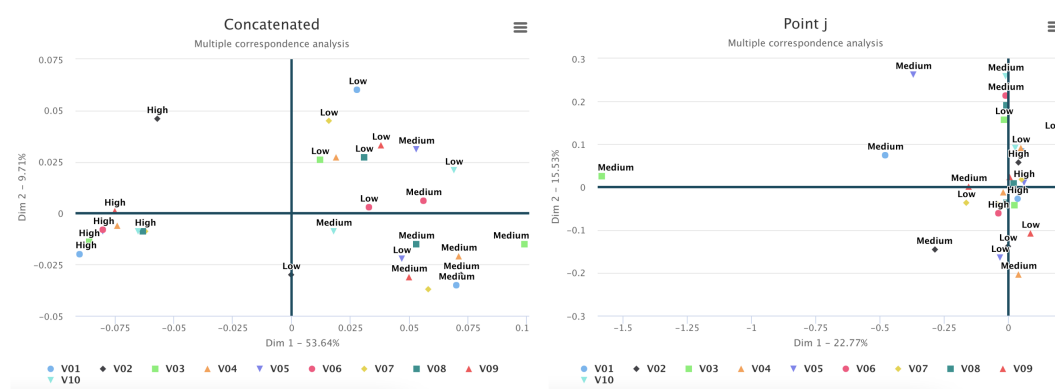


Figure 7. Gráfico de control multivariante T2 Hotelling aplicable a variables cualitativas, *Datak10Contaminated*.

La figura 7 presenta la distribución de las observaciones de las tablas concatenada y j mediante gráficos del MCA. El gráfico de la tabla concatenada, que sirve de referente en control, ya se analizó en la Figura 4. El gráfico de la tabla j muestra una tendencia de los valores medios de las variables a ubicarse al lado izquierdo, alejándose del centro del gráfico, lo que indica que los valores medios

son poco frecuentes. Especial atención merece la variable 3, que registra una observación para el nivel medio con el valor más alejado del grupo. Por el contrario, las categorías altas se han situado en el centro, lo que significa que son muy frecuentes.

Al comparar los gráficos es evidente que la distribución de los datos en el gráfico de la tabla j es diferente de las distribuciones de las demás tablas, y en especial, es diferente de la distribución de los datos en el gráfico de la tabla concatenada, lo que explica por qué el punto j ha sido identificado como fuera de control en el gráfico T2Qv. Esta diferencia se explica en la Tabla 6, que muestra la distancia Chi cuadrado entre las observaciones de la tabla concatenada y la tabla j .

Variables	ChiSq
V1	0.06968
V2	0.05010
V3	0.07601
V4	0.04982
V5	0.05205
V6	0.05603
V7	0.03713
V8	0.03702
V9	0.04395
V10	0.06179

Table 6. Distancia Chi cuadrado entre las masas de columna de la tabla k y la concatenada, *Datak10Contaminated*.

Otra manera de visualizar esta información es a través de un gráfico de barras que genera el aplicativo T2Qv (figura 8).

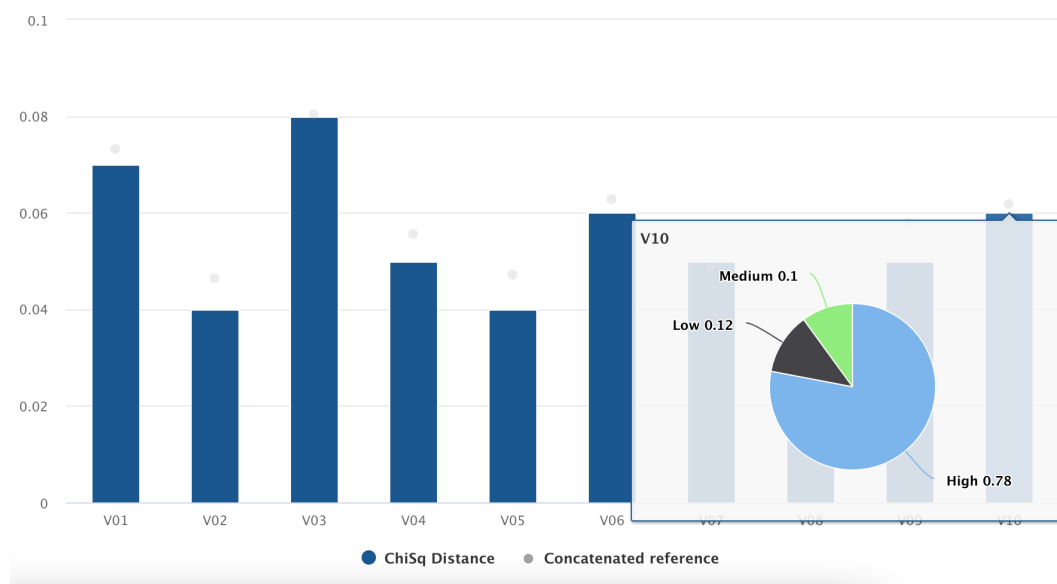


Figure 8. Distancia Chi cuadrado entre las masas de la tabla concatenada y las k tablas, *Datak10Contaminated*.

El gráfico de barras de la figura 8, expresa también la distancia χ^2 entre las masas de la tabla concatenada y las de las k tablas de la base de datos *Datak10Contaminated*, en este caso la j . En la Tabla 6 se observa que las variables V03, V01 y V06 manifiestan las mayores distancias Chi cuadrado entre las masas de la tabla concatenada y la tabla j (0.07700, 0.06968, 0.05938), lo que en la figura 8 se representa con las barras más altas.

La interactividad de este gráfico facilita la observación de la distribución de las categorías de las variables de la tabla analizada, y su comparación con la distribución de las categorías de las variables en la tabla concatenada, como se observa en la figura 9.



Figure 9. Distribución de las categorías de las variables V03, V01 y V06 en la tabla Concatenada y la tabla j en el aplicativo T2Qv.

La figura 9 presenta, en gráficos circulares, la distribución de las categorías de las variables V03, V01 y V06, que registraron mayores distancias Chi cuadrado entre las masas de la tabla concatenada y la tabla j . Los gráficos que corresponden a la tabla concatenada presentan sectores con áreas equivalentes entre sí, lo que se explica por la distribución uniforme de las variables, mientras que, los de la tabla j muestran áreas con tamaños variados, donde la categoría *High* tiene una frecuencia relativa alta en los

tres casos, y *Low*, baja frecuencia. Al comparar estos gráficos se hace evidente que la distribución de las categorías presenta grandes diferencias entre la tabla concatenada y la tabla j . El comportamiento de estas variables tiene mayor incidencia en el desplazamiento de la tendencia central del proceso que, al final, lo lleva a un estado fuera de control. Sin embargo, al tratarse de un contexto multivariante, todas las variables contribuyen en mayor o menor medida a explicar el comportamiento del proceso, de manera que la salida de control no se puede atribuir a la acción individual de una variable, o a la acción por separado de un grupo de ellas, sino al efecto combinado de las variables correlacionadas.

5. Análisis de sensibilidad

Como se ha mencionado, en el gráfico T2Qv un punto fuera de control se interpreta como una tabla (k_i) que incluye una cantidad o una proporción de variables contaminadas. En estos casos, se espera que los puntos en el gráfico T2Qv generalicen el comportamiento de estas diferencias en su distribución y así se supere el límite de control superior (UCL). La ubicación de este límite de control varía en función del número de dimensiones que se representen, así, cuando es alto se logra un desempeño óptimo, mientras que, se introduce inestabilidad y se pierde confiabilidad en los resultados al disminuir el número de dimensiones que se pueden representar.

El gráfico de control propuesto es capaz de detectar un punto fuera de control, aún con un bajo número de variables contaminadas, cuando se trabaja con un alto número de dimensiones. Se recomienda $p - 1$, tal que p es el número total de dimensiones de la matriz inicial (Tabla ??). Cuando se disminuye el número de dimensiones también disminuye la altura del límite de control superior (UCL), en consecuencia, se incrementa el número de puntos fuera de control, aunque no necesariamente las variables expresen diferencias significativas en su valores, crece la probabilidad de obtener error tipo I.

Por consiguiente, la pregunta que surge es hasta cuántas dimensiones se puede disminuir en el análisis sin perder confiabilidad en el resultado. La importancia de esta pregunta radica en la necesidad de disponer un gráfico confiable, que identifique puntos fuera de control aún si se ha aplicado a los datos una técnica de una reducción de dimensiones, sin caer en casos de falso positivo.

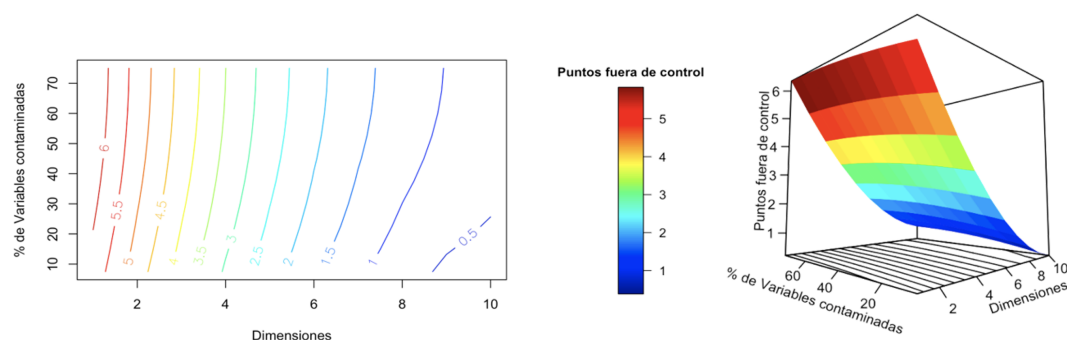


Figure 10. Curvas de nivel y superficie de respuesta obtenidas con el gráfico T2Qv.

El análisis de sensibilidad utiliza curvas de nivel y superficies de respuesta (figura 10) para representar el número de puntos fuera de control, considerando el porcentaje de variables contaminadas de la k_i tabla y el número de dimensiones representadas. Los datos de prueba utilizados en el modelo se registran en 10 tablas, cada una de ellas incluye 10 variables y cada variable tiene tres categorías: *High*, *Medium* y *Low*. La tabla 10 (o tabla j) tiene una distribución diferente de las demás, esta es la tabla contaminada.

Se observa que el modelo es capaz de identificar un punto fuera de control trabajando con $p - 1$ dimensiones (9), aún con un porcentaje bajo de variables contaminadas. Cuando el número de dimensiones disminuye a $p - 2$ (8) y el porcentaje de variables contaminadas es cercano a 100%, detecta correctamente 1 punto fuera de control. Se observa además que cuando el número de dimensiones es menor se pierde estabilidad y se mitiga la potencia de la prueba. En consecuencia, el análisis de

sensibilidad ratifica que el gráfico de control T2Qv tiene un buen rendimiento cuando trabaja con altas dimensiones.

6. Discusión

En el control estadístico de procesos todavía no son muchas las propuestas publicadas sobre gráficos de control para variables cualitativas. Las diferencias entre procedimientos para la determinación de los estadísticos y los gráficos de control en este campo hacen difícil su comparación.

El gráfico de control T2Qv, que se presenta en este artículo, aplica MCA, técnica de análisis multivariante que identifica estructuras latentes que subyacen en el conjunto de datos cualitativos y que involucra una reducción de dimensiones, en consecuencia, desde el comienzo se requiere una tabla de datos con p variables ($p > 3$) dicotómicas o politómicas. Vale recordar que el análisis de sensibilidad determinó que esta propuesta tiene un buen rendimiento cuando trabaja con altas dimensiones y que a bajas dimensiones pierde estabilidad. De esta manera, mientras que en el ejemplo con datos simulados *Datak10Contaminated* que se presenta en esta investigación, el T2Qv analiza el comportamiento de 10 variables, en varios estudios revisados, los casos de aplicación analizan sólo dos o tres, lo que conduciría a la aplicación de un análisis de correspondencias simple, no múltiple. En consecuencia, estos casos no podrían ser tratados con el T2Qv.

Como ejemplos se señala la Combinación lineal óptima de variables Poisson para el SPC multivariados, de Epprecht *et al.* [39], cuyo caso de aplicación registrado en su publicación analiza dos variables relacionadas con el conteo de defectos en la producción de jarrones de cerámica. El gráfico GMDS de Ali and Aslam [55] fue ejemplificado con un conjunto de datos de telecomunicaciones, tomado de Jiang *et al.* [56], que consta de sólo dos variables. El gráfico de control multivariante, desarrollado por Pastuizaca-Fernández *et al.* [44]), para p características de calidad de atributos correlacionadas, que aplica teoría difusa, hace un análisis de dos tablas tomadas de publicaciones de Taleb [43] y Taleb *et al.* [42], la primera con tres variables relacionadas con la comida congelada, y la segunda, con tres variables sobre la producción de porcelana.

Otra de las características del gráfico T2Qv es que cada muestra es un grupo constituido por un conjunto de individuos, una tabla. El ejemplo de datos simulados *Datak10Contaminated* incluye un conjunto de 10 tablas y 11 variables, cada tabla es una muestra, está formada por 100 observaciones y aparece representada como un punto en el gráfico T2 de Hotelling.

En publicaciones de varios autores se puede constatar que en sus ejemplos de aplicación se analiza una sola tabla, de dimensiones n (filas) $\times p$ (variables), donde cada n_i fila es una muestra. Por ejemplo, el gráfico de control MNP, de Lu [40] contiene en su artículo una tabla de datos simulados de 30 muestras, donde cada una de ellas es un único individuo (objeto) que registra el conteo de defectos para tres características de la calidad. Asimismo, la ejemplificación que Chiu and Kuo [36] presentaron de su gráfico de control MP se hizo con una tabla de datos simulados de 26 muestras, donde cada muestra representa a un individuo al que se le registra el D número de defectos o no conformidades asociadas a tres características de calidad.

En el gráfico de control T2Qv que se presenta en este artículo, cada uno de los individuos (filas) que conforman las diferentes muestras pueden tener distintas configuraciones en función de las categorías de las variables.

Por el contrario, otros autores que han investigado sobre gráficos de control multivariante para datos de atributos, aunque en su análisis consideran varias características de calidad, al final clasifican a cada individuo por una sola de las variables analizadas. Es el caso de Ranjan-Mukhopadhyay [41], cuya propuesta se demuestra con un caso de aplicación que controla 7 características de calidad en 24 muestras cuyo tamaño varía entre 20 y 404 individuos. Las variables responden a 6 tipos de defectos de la pintura en la cubierta de ventiladores de techo: cobertura deficiente, desbordamiento, defecto de empanada, burbujas, defectos de pintura, defectos de pulido. La séptima característica es la ausencia de defectos. A cada individuo se lo clasifica por su defecto más predominante, por consiguiente, en

su registro sólo aparece un tipo de defecto o ausencia de defectos, lo que resulta en una pérdida de información sobre el efecto combinado de las variables sobre el proceso.

Los gráficos de control que incorporan técnicas multivariantes dirigen su aplicación al análisis de variables categóricas y numéricas en una misma tabla de datos. Por ejemplo, el gráfico de control PCA Mix [30] convierte las variables de atributo en ficticias y las trata junto con las variables continuas, para generar una matriz Kernel y calcular las componentes principales. Una debilidad de esta propuesta está en que su rendimiento disminuye en presencia de proporciones extremadamente desequilibradas de las categorías de variables cualitativas, situación que se propone corregir aplicando el método Kernel PCA (KPCA), una versión no lineal del PCA convencional que modela datos de distribuciones no gaussianas.

Estos ajustes al rendimiento corresponden a una fase II de los gráficos de control, con miras a su optimización. El T2Qv es un gráfico de control de procesos multivariantes que maneja variables cualitativas en fase I, en consecuencia, no se ha considerado aún la evaluación de su eficiencia, por lo que ambos gráficos no son comparables. Sin embargo, es precisamente el cambio en la distribución de las categorías de las variables, de balanceada a desequilibrada, o viceversa, lo que detecta el T2Qv como punto fuera de control.

Una propuesta interesante es la de Saltos Segura *et al.* [45], quienes utilizan el concepto de profundidad para el análisis de una tabla de datos en el campo de la educación, pero, al final la representación del rendimiento académico es univariante y se realiza mediante una carta de control \bar{p} . El ACM, siendo una técnica factorial que trabaja en términos de asociación de variabilidad, hace que la información común a todos los casos, o a la gran mayoría, sea estable y por lo tanto no aparezca en los ejes primarios, en el mejor de los casos se concentra en el origen del gráfico. De hecho en el aplicativo T2Qv, cuando una categoría de alguna variable es constante en una tabla de la base de datos, se reporta un error que impide la ejecución, pero, si hay por lo menos un caso diferente, éste se representaría como un punto muy alejado en algún extremo del gráfico, mientras que la categoría que registre la mayor frecuencia se ubicaría como un punto en el centro. Esto podría constituir una característica propia de la naturaleza del ACM, que sólo puede representar parcialmente la variabilidad de la información en sus dos dimensiones.

Para corregir este problema, la metodología propuesta en esta investigación, aunada a la técnica del ACM, utiliza el gráfico T2Qv que, como queda establecido en el análisis de sensibilidad, funciona mejor con la mayor cantidad de dimensiones, es decir, recoge la mayor cantidad de variabilidad para identificar el punto (la tabla punto) fuera de control. Además, un análisis comparativo posterior establece la distancia χ^2 entre los valores reportados por las categorías de la tabla concatenada y la tabla punto para una variable específica y la representa en un gráfico de barras, lo que permite la identificación de las variables que más están produciendo anomalías. Finalmente, el aplicativo facilita la comparación de la distribución de las categorías de la variable analizada en la tabla punto y la concatenada a través de un recurso gráfico interactivo que utiliza gráficos de sectores.

Una oportunidad para futuras investigaciones relacionadas con el control multivariante para variables cualitativas podría ser la optimización del gráfico, con un límite de control que se ajuste a los parámetros específicos del análisis, llevándolo hasta a una fase II del control estadístico de procesos. Otra oportunidad sería el desarrollo de una metodología que vaya más allá del análisis de la primera dimensión latente, que es lo que hace la propuesta de esta investigación cuando aplica ACM. Podría ser viable la incorporación, por ejemplo, de un Meta Biplot [20], técnica que analiza de forma cruzada todas las dimensiones latentes.

7. Conclusiones

En este artículo se ha presentado una herramienta para control estadístico de procesos multivariantes que realiza un análisis de los datos cualitativos, al que se ha denominado Gráfico de control T2Qv, basado en un Análisis de Correspondencias Múltiples. Las coordenadas normalizadas, se representan mediante el gráfico T2 de Hotelling robusto.

Para facilitar la difusión y aplicación del método propuesto, se ha desarrollado un paquete estadístico computacional reproducible en R, denominado T2Qv y disponible en CRAN, que permite visualizar los resultados de forma plana o interactiva, además, presenta un panel Shiny que contiene todas las funciones integradas en un mismo espacio.

Esta propuesta genera un gráfico del MCA de la tabla concatenada, que sirve de referente para comparar otros gráficos del MCA de las tablas que hayan sido identificadas como puntos fuera de control en el gráfico de Hotelling. Allí se puede verificar qué categorías de las variables han tenido variaciones en su ubicación en ambos gráficos, que pueden estar provocando cambios en la centralidad del proceso y ocasionando el estado de fuera de control.

Para facilitar la interpretación del comportamiento de las variables se realiza un análisis de la distancia Chi cuadrado entre las categorías de la tabla concatenada y de las tablas reportadas como fuera de control, analítica y gráficamente, incluyendo gráficos interactivos que presentan la distribución porcentual de las categorías de las variables analizadas.

En un contexto multivariante, todas las variables contribuyen en mayor o menor medida a explicar el comportamiento del proceso, de manera que la salida de control no se puede atribuir a la acción individual de una variable, o a la acción por separado de un grupo de ellas, sino al efecto combinado de variables correlacionadas. Por eso es preciso el enfoque multivariante en el control de procesos estadísticos.

El análisis de sensibilidad determinó que el gráfico de control T2Qv tiene un buen rendimiento cuando trabaja con altas dimensiones, pero, pierde estabilidad a bajas dimensiones.

En el control estadístico de procesos multivariantes para variables cualitativas todavía no son muchas las propuestas publicadas. Las diferencias entre procedimientos para la determinación de los estadísticos y los gráficos de control en este campo hacen difícil su comparación.

El gráfico de control T2Qv atiende la necesidad de un gráfico multivariante de control estadístico para variables cualitativas en procesos sociales, donde es muy común el uso de variables nominales y ordinales.

References

- Gutiérrez, H.; de la Vara Salazar, R. *Control estadístico de la calidad y seis sigma*; Vol. 3, McGraw Hill Education, 2013; p. 152 – 253.
- Ramos, M. Una alternativa a los métodos clásicos de control de procesos basada en coordenadas paralelas, métodos Biplot y Statis. PhD thesis, 2017.
- Li, J.; Tsung, F.; Zou, C. Directional control schemes for multivariate categorical processes. *Journal of Quality Technology* **2012**, *44*, 136–154.
- Hotelling, H. Multivariate quality control. *Techniques of statistical analysis*. McGraw-Hill, New York **1947**.
- Lowry, C.A.; Woodall, W.H.; Champ, C.W.; Rigdon, S.E. A multivariate exponentially weighted moving average control chart. *Technometrics* **1992**, *34*, 46–53.
- Roberts, S. Control chart tests based on geometric moving averages. *Technometrics* **2000**, *42*, 97–101.
- Crosier, R.B. Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics* **1988**, *30*, 291–303.
- Page, E. Continuous inspection schemes. *Biometrika* **1954**, *41*, 100–115.
- APARISI, F. Hotelling's T2 control chart with adaptive sample sizes. *International Journal of Production Research* **1996**, *34*, 2853–2862, [<https://doi.org/10.1080/00207549608905062>]. doi:10.1080/00207549608905062.
- Aparisi, F.; Haro, C.L. Hotelling's T2 control chart with variable sampling intervals. *International Journal of Production Research* **2001**, *39*, 3127–3140, [<https://doi.org/10.1080/00207540110054597>]. doi:10.1080/00207540110054597.
- Faraz, A.; Parsian, A. Hotelling's T2 control chart with double warning lines. *Statistical Papers* **2006**, *47*, 569–593. doi:10.1007/s00362-006-0307-x.

12. Ruiz-Barzola, O. Gráficos de Control de Calidad Multivariantes con Dimension Variable. PhD thesis, Universitat Politècnica de València, 2013.
13. Shabbak, A.; Midi, H. An improvement of the hotelling statistic in monitoring multivariate quality characteristics. *Mathematical Problems in Engineering* **2012**, 2012.
14. Liu, Y.; Liu, Y.; Jung, U. Nonparametric multivariate control chart based on density-sensitive novelty weight for non-normal processes. *Quality Technology & Quantitative Management* **2020**, 17, 203–215.
15. Xue, L.; Qiu, P. A nonparametric CUSUM chart for monitoring multivariate serially correlated processes. *Journal of Quality Technology* **2020**, pp. 1–14.
16. Mahalanobis, P. On the generalised distance in statistics. Proceedings of the national Institute of Science of India, 1936, Vol. 12, pp. 49–55.
17. Gower distance-based multivariate control charts for a mixture of continuous and categorical variables. *Expert Systems with Applications* **2014**, 41, 1701–1707. doi:10.1016/j.eswa.2013.08.068.
18. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science* **1901**, 2, 559–572.
19. Gabriel, K.R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **1971**, 58, 453–467.
20. Galindo-Villardón, P.; Vicente-Villardón, J.; Zarza, C.A.; Fernandez-Gómez, M.J.; Martín, J. JK-META-BIPLLOT: una alternativa al método STATIS para el estudio espacio temporal de ecosistemas.
21. Benzecri, J. OL'analyse des correspondances. En *L'Analyse des Données: Leçons sur L'analyse Factorielle et la Reconnaissance des Formes et Travaux*; Paris - 1973, 1973.
22. des Plantes, L. Structuration des tableaux à trois indices de la statistique. *Université de Montpellier II, Thesis* **1976**.
23. Robert, P.; Escoufier, Y. A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Journal of the Royal Statistical Society Series C: Applied Statistics* **1976**, 25, 257–265.
24. Lavit, C.; others. Présentation de la méthode STATIS permettant l'analyse conjointe de plusieurs tableaux de données quantitatives. *Les Cahiers de la Recherche Développement* **1988**, pp. 49–60.
25. Inselberg, B.; Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. Proceedings of the first IEEE conference on visualization: visualization90. IEEE, 1990, pp. 361–378.
26. Edwards, A.; Cavalli-Sforza, L. A method for cluster analysis. *Biometrics* **1965**, pp. 362–375.
27. Filho, D.; de Oliveira, L. Multivariate quality control of batch processes using STATIS. *The International Journal of Advanced Manufacturing Technology* **2016**, 82, 867–875.
28. Ramos-Barberán, M.; Hinojosa-Ramos, M.V.; Ascencio-Moreno, J.; Vera, F.; Ruiz-Barzola, O.; Galindo-Villardón, M.P. Batch process control and monitoring: a Dual STATIS and Parallel Coordinates (DS-PC) approach. *Production & Manufacturing Research* **2018**, 6, 470–493, [<https://doi.org/10.1080/21693277.2018.1547228>]. doi:10.1080/21693277.2018.1547228.
29. Ahsan, M.; Mashuri, M.; Kuswanto, H.; Prastyo, D.D.; Khusna, H. Multivariate control chart based on PCA mix for variable and attribute quality characteristics. *Production & Manufacturing Research* **2018**, 6, 364–384, [<https://doi.org/10.1080/21693277.2018.1517055>]. doi:10.1080/21693277.2018.1517055.
30. Ahsan, M.; Mashuri, M.; Wibawati.; Khusna, H.; Lee, M.H. Multivariate Control Chart Based on Kernel PCA for Monitoring Mixed Variable and Attribute Quality Characteristics. *Symmetry* **2020**, 12, doi:10.3390/sym12111838.
31. Ahsan, M.; Mashuri, M.; Khusna, H. Comparing the performance of Kernel PCA Mix Chart with PCA Mix Chart for monitoring mixed quality characteristics. *Scientific Reports* **2022**, 12, 1–12.
32. Ahsan, M.; Mashuri, M.; Kuswanto, H.; Prastyo, D.D.; Khusna, H. Outlier detection using PCA mix based T2 control chart for continuous and categorical data. *Communications in Statistics - Simulation and Computation* **2021**, 50, 1496–1523, [<https://doi.org/10.1080/03610918.2019.1586921>]. doi:10.1080/03610918.2019.1586921.
33. Farokhnia, M.; Niaki, S.T.A. Principal component analysis-based control charts using support vector machines for multivariate non-normal distributions. *Communications in Statistics - Simulation and Computation* **2020**, 49, 1815–1838, [<https://doi.org/10.1080/03610918.2018.1506032>]. doi:10.1080/03610918.2018.1506032.
34. Liu, Y.; Liu, Y.; Jung, U. Nonparametric multivariate control chart based on density-sensitive novelty weight for non-normal processes. *Quality Technology & Quantitative Management* **2020**, 17, 203–215.

35. Holgate, P. Estimation for the bivariate Poisson distribution. *Biometrika* **1964**, *51*, 241–287.
36. Chiu, J.E.; Kuo, T.I. Attribute control chart for multivariate Poisson distribution. *Communications in Statistics-Theory and Methods* **2007**, *37*, 146–158.
37. Lee, L.H.; Costa, A.F.B. Control charts for individual observations of a bivariate Poisson process. *The International Journal of Advanced Manufacturing Technology* **2009**, *43*, 744–755.
38. Laungrungrong, B.; M, C.B.; Montgomery, D.C. EWMA control charts for multivariate Poisson-distributed data. *International Journal of Quality Engineering and Technology* **2011**, *2*, 185–211.
39. Epprecht, E.K.; Aparisi, F.; García-Bustos, S. Optimal linear combination of Poisson variables for multivariate statistical process control. *Computers & operations research* **2013**, *40*, 3021–3032.
40. Lu, X. Control chart for multivariate attribute processes. *International Journal of Production Research* **1998**, *36*, 3477–3489.
41. Ranjan-Mukhopadhyay, A. Multivariate attribute control chart using Mahalanobis D 2 statistic. *Journal of Applied Statistics* **2008**, *35*, 421–429.
42. Taleb, H.; Limam, M.; Hirota, K. Multivariate fuzzy multinomial control charts. *Quality Technology & Quantitative Management* **2006**, *3*, 437–453.
43. Taleb, H. Control charts applications for multivariate attribute processes. *Computers & Industrial Engineering* **2009**, *56*, 399–410.
44. Pastuizaca-Fernández, M.N.; Carrión-García, A.; Ruiz-Barzola, O. Multivariate multinomial T 2 control chart using fuzzy approach. *International Journal of Production Research* **2015**, *53*, 2225–2238.
45. Saltos Segura, G.; Flores Sánchez, M.; Horna Huaraca, L.; Morales Quinga, K. NEW METHODOLOGIES APPLIED TO MULTIVARIATE MONITORING OF STUDENT PERFORMANCE USING CONTROL CHARTS AND THRESHOLD SYSTEMS. *Perfiles* **2020**, *1*, 68–74.
46. López, C.P. *Técnicas de análisis multivariante de datos*; Pearson Educación, 2004.
47. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series* **1901**, *2*, 417 – 441. doi:10.1080/14786440109462720.
48. Hotelling, H. Analysis of a complex of statistical variables into principal components. **1933**. *24*, 417 – 441. doi:10.1037/h0071325.
49. Ch, S.; others. General intelligence objectively determined and measured. *American Journal of Psychology* **1904**, *15*, 201–293.
50. Thurstone, L.L. Multiple-factor analysis; a development and expansion of The Vectors of Mind. **1947**.
51. Kaiser, H. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **1958**, *23*, 187–200.
52. Michailidis, G.; Leeuw, J.D. The Gifi system of descriptive multivariate analysis. *Statistical Science* **1998**, pp. 307–336.
53. Escofier, B.; Pagès, J. Multiple factor analysis (AFMULT package). *Computational Statistics & Data Analysis* **1994**, *18*, 121–140. doi:https://doi.org/10.1016/0167-9473(94)90135-X.
54. Rojas-Preciado, W.; Rojas-Campuzano, M.; Galindo-Villardón, P.; Ruiz-Barzola, O. *T2Qv: Control Qualitative Variables*. R package version 0.1.0.
55. Ali, M.R.; Aslam, M. Design of control charts for multivariate Poisson distribution using generalized multiple dependent state sampling. *Quality Technology & Quantitative Management* **2019**, *16*, 629–650.
56. Jiang, W.; Au, S.; Tsui, K.L.; Xie, M. Process monitoring with univariate and multivariate c-charts. *Technical Report, the Logistics Institute, Georgia Tech, and the Logistics Institute-Asia Pacific* **2002**.